Postal Regulatory Commission
Submitted 12/8/2020 3:22:02 PM
Filing ID: 115289
Accepted 12/8/2020

Reply Report of A. Thomas Bozzo
In Response to Comments of the Public Representative
Docket No. RM2020-13
December 8, 2020

Biographical Sketch

A. Thomas Bozzo is a Vice President with Christensen Associates, where he has been employed since 1996. Dr. Bozzo has a Bachelor of Arts degree in economics and English from the University of Delaware, and a Ph.D. in economics from the University of Maryland-College Park. His areas of expertise include economic cost measurement; postal, railroad, and telecommunications regulation; and applied econometrics and statistics. He currently leads Christensen Associates' area of practice responsible for clerk and mail handler cost and labor productivity production for the Postal Service's Cost and Revenue Analysis and Annual Compliance Reports. Dr. Bozzo has also been involved with numerous other projects for the Postal Service, focusing on applications of econometrics, sample-based data, and economic cost theory for measurement of costs and productivity for postal activities and products, including Total Factor Productivity and surveys of field operations. Dr. Bozzo has presented testimony on mail processing costs in Dockets No. R2000-1, R2005-1, and R2006-1, and testified on the In-Office Cost System (IOCS) survey instrument design in Docket No. R2006-1. He has consulted for USPS OIG on subjects including a review of Postal Service demand models (Report No. RARC-WP-13-008, May 1, 2013) and an assessment of costs related to provision of service standards. He was a primary author of the Christensen Associates 2008 study of freight railroad competition for the Surface Transportation Board, for which he led the study's econometric analysis of the determinants of rail freight pricing. Dr. Bozzo has also been involved in projects in other practice areas, including studies of electricity and natural gas demands, econometric analyses of energy efficiency programs, analysis of telecommunications cost models for projects related to universal service proceedings, and a variety of litigation support projects.

Purpose of Reply Report

This report responds to criticisms of Proposal Six leveled in the Public Representative's Docket No. RM2020-13 comments, as revised on November 30, 2020 ("PR Comments"). I address three main issues raised by the Public Representative.

First, while the Public Representative admits that volume changes since the last Postal Reorganization Act are not the only motivations for Proposal Six, he pays scant attention to key differences between the current mail processing environment and that of 15 (or more) years ago that were the subject of major controversies in the past mail processing variability studies. I discuss key context related to the transition away from more complex mailflows among relatively immature automated operations at the time of the earlier studies, the greatly

diminished role of manual processing in the current system, and the reliable data collection methods employed to collect mail processing volume data in MODS for the operations that are the subjects of Proposal Six. These factors justify the relatively simple models in Proposal Six.

Second, I describe significant errors in the Public Representative's discussion of current-system operations, which misread the Variability Report to suggest, incorrectly, that technological changes in flats operations should have increased staffing flexibility for the operations. In addition, I describe a fatal error in the Public Representative's analysis that purports to show trends in machine inventories, caused by the Public Representative's misinterpretation of observation counts in the MODS dataset for machine counts. As a result of the error, the Public Representative fails to note the existence of constraints in adjusting machine inventories that are more binding for letter equipment than flat equipment.

Third, I discuss the Public Representative's critiques of various details of the econometric methodology in Proposal Six. In a number of cases, the Public Representative omits or ignores justifications for Proposal Six methods discussed in the Variability Report and subsequent responses to Chairman's Information Requests. In addition, I describe another fatal error in the Public Representative's accompanying analysis of the data screening methods used in Proposal Six: rather than examining observations from the tails of the distribution of labor productivity (TPF/workhour, the screening criterion), he extracts observations from the tails of the distributions of workhours themselves. As a result, he mistakenly contends that the Proposal Six analysis improperly omits presumptively valid observations that actually are included. I also show that the Proposal Six results subject to the critiques are justifiable and/or robust to certain alternative specifications, notably substituting MODS FHP for TPF as the sorting output measure.

I. Changes to letter and flat sorting operations since Docket No. R2006-1 justify reassessing the criteria enumerated by the Public Representative

The Public Representative's comments fault the Postal Service for advancing Proposal Six without comprehensively addressing a number of methodological criteria from the Commission's rejection of previous mail processing variability models in Docket No. R2006-1 and previous rate cases prior to the enactment of the Postal Accountability and Enhancement Act. PR Comments at 1-2; 9-10. The Public Representative gives scant consideration to whether present operational circumstances warrant a reassessment or whether using available, if perhaps imperfect data—as opposed to ideal but nonexistent data such as hypothetical plant-level RPW-type volumes—to establish an empirical basis for the volume-variability would improve the overall quality of the costing data employed by the Commission.

In fact, many of the concerns raised with respect to modeling sorting operations revolve around more complex mailflows with greater substitution possibilities among operations and much larger costs in manual operations. The three MODS operation groups that are the subject of the Proposal Six analysis now represent the vast majority of costs in the letter and flat mailstreams

and have limited substitutability with other operations. The MODS data on processing volumes and machine utilization for those operations are generated by reliable automated data acquisition systems, unlike workloads used in the rate case studies that were inherently subject to measurement error from data conversions.

a. Cost structure changes in both letter and flat sorting operations justify the use of simpler models

The Public Representative notes that declines in letter and flat volumes are not the only justifications for Proposal Six (PR Comments at 10-11) but fails to consider the implications of changes in letter and flat sorting technologies and cost structure for modeling the operations. The mailflows considered in the analyses from Dockets No. R97-1 to R2006-1 were more complex, generally involving multiple major types of processing equipment and much greater reliance on manual processing in both absolute and relative terms.

Table 1, below, shows that in the present system, automated mail processing is more highly concentrated in one major type of equipment for each of the letter and flat mailstreams—the DBCS for letter sorting, and the AFSM 100 for flat sorting. While FSS operations comprise a substantial fraction of flat distribution costs, the FSS is not generally substitutable for AFSM 100 processing, as its sole processing function is automated delivery point sequencing of flats, while the AFSM 100 is not technically capable of carrying out flats DPS and is not substitutable for FSS. Manual sorting operations have fallen from over 40 percent of costs in both the letter and flat mailstreams in the FY1996 base year for Docket No. R97-1 to less than 15 percent in FY2019.

Table 1. Composition of Letter and Flat Sorting Labor Costs, Selected Fiscal Years

	FY1996 (R97-1)		FY2005 (R2006-1)		FY2019 (ACR2019)	
Letter Sorting	Cost	Share	Cost	Share	Cost	Share
DBCS (w/ MPBCS)	681,360	23%	1,482,016	57%	1,511,307	86%
OCR	224,198	8%	201,547	8%	0	0%
LSM	731,680	25%	0	0%	0	0%
Manual Letters	1,342,326	45%	917,249	35%	243,815	14%
Total Letters	2,979,564		2,600,812		1,755,122	

	FY1996 (R97-1)		FY2005 (R2006-1)		FY2019 (ACR2019)	
Flat Sorting	Cost	Share	Cost	Share	Cost	Share
FSM (775, 881)	736,969	59%	0	0%	0	0%
FSM 1000	0	0%	218,122	17%	0	0%
AFSM 100 w/ Prep	0	0%	837,742	65%	604,973	65%
FSS	0	0%	0	0%	210,432	23%
Manual Flats	514,848	41%	239,251	18%	112,982	12%
Total Flats	1,251,817		1,295,115		928,387	

Source
FY1996: Docket No. R97-1 USPS LR-H-146
FY2005: Docket No. R2006-1, USPS-LR-L-55, LR-L-55 pt1.xls
FY2019: USPS-RM2020-13-1, FY07-19 MP Costs w-RPW_v.xlsx

Insofar as letter and flat sorting costs are highly concentrated in incoming automated operations with no significant substitution possibilities, modeling DBCS, AFSM 100, and FSS operations individually is a choice consistent with the structure of current operations. The operation groups are distinguishable by the primary shape of mail processed therein. Given their limited role in mail processing plants, Proposal Six conservatively maintains current Commission methodology for manual variabilities.

As the Postal Service has noted, outgoing processing also represents relatively small and declining shares of current-system sorting operations for both letters and flats, as a consequence of relatively more rapid volume declines in full-network volumes such as single-piece mail. Combining relatively small and faster-declining outgoing operations with much larger incoming operations is also reasonable. Distinguishing them where practicable does not materially affect estimated variabilities. See Response to ChIR No. 2, question 5(e).

b. Automated sorting volumes are generated by methods that involve negligible measurement error, and this largely obviates the need for instrumental variables estimation

The Public Representative also cites past concerns regarding measurement error in MODS data without consideration of issues that were specific to the past rate case studies, notably the measurement of manual letter and flat sorting volumes and of MODS First Handling Pieces (FHP) in all operations. Both manual volumes and FHP measurements at the time relied on estimation procedures such as conversion of weights to pieces that were inherently subject to measurement error. Aggregating data over time or over operations within a cost pool does not eliminate the error component—and more specifically, the measurement error variance is central to the errors-in-variables issue for regression modeling.[1] The use of instrumental variables had been pursued in the rate case analyses primarily to address the errors-in-variables issues inherent in manual workloads and FHP.

Errors-in-variables is not an issue for Proposal Six because the MODS sorting volumes employed in the analysis are based on direct piece counts from the equipment that are automatically transmitted to MODS via the webEOR system along with other machine operating statistics such as machine runtime. Response to ChIR No. 3, question 2(a). The workloads used as independent variables in the models are accurately measured in principle, and the Public Representative makes no showing that machine-derived volumes and other statistics in MODS are materially inaccurate in any way. The absence of comparably reliable workload data for manual letter and flat operations, however, is a factor that potentially may justify the implicit retention of variabilities from currently accepted methods for those operations in Proposal Six.

It is important to keep in mind that the errors-in-variables problem applies specifically to independent (explanatory) variables in regressions. Errors in dependent variables such as workhours do not themselves cause the estimated coefficients to be biased or inconsistent, as measurement errors in dependent variables are subsumed in the regression models' error terms.[2]

Indeed, the Postal Service also improved the measurement of FHP for automated letter and flat operations subsequent to Docket No. R2006-1 by switching from weight conversion to EOR data analytics. The current methodology is not exactly a direct count of first handlings. The system uses mailflows within plants to estimate subsequent handlings at the level of mail processing runs and subtracts subsequent handlings from Total Piece Handlings (TPH)—which are directly counted—to obtain FHP. Nevertheless, the current FHP method is much closer to a direct count

[1] See, e.g., Cheng Hsiao, *Analysis of Panel Data* (Cambridge University Press, 1986) at 63-65.

[2] Measurement errors may affect the efficiency of estimation by increasing the residual variance relative to a case of dependent variables measured without error. Also, it may be necessary to control appropriately for systematic errors, which contributes to the justification for using techniques such as fixed-effects estimators.

for automated operations than the former weight-conversion approach. As discussed below, FHP is not preferred as the sorting volume or output measure for other reasons.

II. The Public Representative's characterizations of sorting operations are marked with a number of basic errors that lead to mischaracterizations of operational constraints facing the Postal Service

In addition to failing to consider whether the structure of Proposal Six is justifiable by the present configuration of sorting operations, the Public Representative makes a number of errors in his account of sorting operations that lead him to downplay constraints on labor flexibility faced by the Postal Service that differentially affect letter and flat operations and explain the pattern of results in Proposal Six.

a. The Public Representative's characterization of staffing flexibility issues with AFSM 100 is backwards

A core claim of the Public Representative is that declining productivities in flat operations are largely a phenomenon of operational management decisions, rather than being consequences of technologies used in the mail processing system. PR Comments at 18-19. The Proposal Six analysis does not, in fact, claim that technological factors are the only forces at play in determining cost levels and variability; rather, both factors may contribute at least to some extent. Variability Report at 5; MPA, et al., Comments at 3-4.[3]

The Public Representative overlooks technical and operational factors that contribute to relative cost inflexibility for flats due to a number of erroneous characterizations of the operations under study in Proposal Six. An elementary, if relatively inconsequential, error is an incorrect statement that DBCS staffing is one clerk per machine (PR Comments at 11), when the Variability Report (citing operations testimony from Docket No. R2006-1, USPS-T-42) clearly states that the standard complement is two clerks: one feeder and one sweeper. Variability Report at 5.

A more significant error is an apparent confusion regarding the staffing requirements for AFSM 100 equipment with and without automated induction (AI) equipment. The Public Representative incorrectly concludes that AFSM 100 equipment, in its original configuration *without* AI, had a fixed complement, such that managerial limitations on staffing flexibility "more likely refer to older vintage AFSM100s". PR Comments at 11-12. The Public Representative misreads the Variability Report, which states that an AFSM 100 complement of five (not including flat preparation labor) would reflect the machine "in *full* operation."

[3] The Public Representative does not explain why the mere role of management in Postal Service costs should militate against empirical measurement of labor elasticities. Management actions play some role in every cost component using econometric variabilities in approved methods. If costs actually are not perfectly flexible with respect to volume changes, then cost elasticities should reflect that.

Variability Report at 6 (emphasis added). To clarify, non-AI AFSM 100 equipment can be operated with less than a full complement—e.g., one or two feeders rather than three; one sweeper instead of two—at times when mail volumes are light.[4] For AI equipment, one clerk may effectively operate all three feeders, which may be efficient (particularly in high-volume periods) but clearly is relatively inflexible with respect to volume changes.[5] The upgraded AFSM 100 equipment subject to relative staffing inflexibility is the newer-vintage AFSM 100 equipment,[6] which is the largest component of current AFSM 100 costs. Response to ChIR No. 2, question 5(c).

b. The Public Representative's analysis purporting to show trends in machine inventories and machines per site is invalid because the Proposal Six data do not include machine counts

In support of his contention that management actions are responsible for observed productivity changes, the Public Representative presents what he claims is an analysis that purports to show that machines per site were little changed while total machines declined more substantially (PR Comments at 19-22). The analysis, however, is fatally flawed due to a basic misinterpretation of the data: the Public Representative mistakes MODS *observation* counts (i.e., numbers of records in the MODS dataset by 3-digit operation number, month, and finance number, for the AFSM 100, DBCS, and FSS operation groups) for *machine* counts.[7]

There is no simple correspondence between MODS operation numbers and machines. MODS operation numbers collect data from all machines at a facility that run sort schemes associated with the operation. Individual machines also may run sort schemes associated with multiple MODS operations, reporting volumes of mail processed and other statistics under the appropriate operation codes. The result is that the Public Representative's analysis greatly mischaracterizes equipment counts both in total and on a per-site basis. Table 2, below, shows the differences between the Public Representative's counts and the Postal Service's equipment inventory, and also shows the number of machines per site, as of the start of FY2019.

[4] I have observed this directly in visits to Postal Service mail processing facilities. I have also observed clerks monitoring AI feeders in periods where mail volumes were insufficient to maintain utilization of all three induction stations.

[5] The Automated Tray Handling System (ATHS) upgrade similarly reduces maximum staffing requirements for the sweep side of the machine at a potential cost of more limited downward staffing flexibility.

[6] Witness McCrery stated in Docket No. R2006-1, USPS-T-42, that procurement of the upgrades was in process at the time the testimony was written.

[7] The variable in question, obs_count, is generated by a count function that generates the number of observations (by counting instances of the facility finance number) within sites, operation groups, and month. As an example, the Public Representative's analysis shows that most FSS sites have "20-26… machines." PR Comments at 21 (Chart 3). This reflects FSS sites typically reporting two MODS observations per month—one each for MODS operations 530 and 538—leading to 24 annual observations.

Table 2. Equipment counts per site by machine type

	AFSM 100	DBCS	FSS
Machines (FY2019) /1	469	3856	100
Active Sites (FY2019 Avg.) /2	196	209	42
Machines Per Site	2.4	18.5	2.4
PR "Machines" (FY2019) /3	1098	1895	1026
% Difference, PR vs Actual Machines	134%	-51%	926%

Source:
1/ ACR2019, USPS-FY19-8, FCILTY19.xlsx, worksheet "Equipment," Mid-FY deployments; DBCS includes "OCR Replacements" (i.e., DIOSS)
2/ USPS-RM2020-13-1, Fig 3 Facilities.xlsx
3/ PR Comments at 20-21

The Public Representative's error in interpreting the data obscures additional constraints related to automation capacity that are more binding for flat versus letter operations. A consequence of the average number of AFSM 100s per site being 2.4 machines is that many sites have only one or two AFSM 100s, and thus have limited or no ability to reduce flat sorting equipment while retaining sufficient (in some cases, any) automated flat processing capacity.[8] In contrast, the higher average number of DBCS machines per site provides more opportunities to adjust capacity in relatively small increments.

III. The Public Representative's criticisms of Proposal Six econometric methodology details are unjustified

The Public Representative's critiques of the Proposal Six, like his analysis of operational factors, is marked by a number of basic errors of analysis and even errors in characterizing common econometric practice. Several examples are relatively inconsequential in themselves. For example, the Public Representative makes a distinction between "log-linear" and "log-log" functional forms, asserting that it would be problematic to take the log of a dummy variable. PR Comments at 13. Econometric practice is not to transform dummy variables; rather, the variables are interpreted as multiplicative shifters.[9] He incorrectly characterizes the Proposal Six fixed-effects models as having "machine clusters," which appears to conflate the

[8] Note also that equipment requirements for both letters and flats depend in part on delivery network characteristics, since incoming secondary sorting and delivery point sequencing require at least one output stacker or bin per carrier route. These requirements are met in part by running sort programs for different zones sequentially on the same machines, but that is limited by the amount of time in the day, as well as capacity required for other incoming and outgoing processing. Thus, while the presence of a single machine may be a hard integer constraint, particularly for larger facilities two (or more) machines may represent the effective required minimum capacity.

[9] It is possible to implement additive shifters with dummy variables in an exponential model that is non-linear in its parameters.

implementation of site-specific (not machine-specific) effects in the models with the use of clustering (also by site) in Proposal Six as a method of generating standard errors that are robust to heteroskedasticity and within-site correlation of the residuals. PR Comments at 13; see also the response to Chairman's Information Request No. 3, Question 4. In reporting results from the models without lags or seasonal dummies (PR Comments at 15 [Table 3]), the Public Representative reported results from models based on unscreened data rather than estimates comparable to the Proposal Six results (which are reported directly in the Variability Report at 23 [Table 4]). However, these minor if pervasive errors set the stage for more serious errors of analysis that are discussed below.

a. The use of elasticities of workhours with respect to volumes is appropriate

The Public Representative inappropriately implies that Proposal Six should have employed the runtime regression as the "appropriate model" rather than the workhour model for Proposal Six. PR Comments at 24. In doing so, the Public Representative misconstrues both the application of Proposal Six and the roles of the two sets of equations. The goal of Proposal Six is to establish an empirical basis for the volume-variability factors for mail processing labor cost pools, and the MODS runtime variable is a measure of machine utilization, not labor usage. Additionally, the quantity of pieces processed on the machines (MODS TPF) is the appropriate output measure for the operation groups included in Proposal Six. Variability Report at 6-9.

The Variability Report does not present the runtime and workhour models as alternatives, but rather as complements. As explained in the response to ChIR No. 1, question 10, the runtime models are presented to examine whether the machine operating hours (which in turn must be staffed) are proportional to sorting volumes for the Proposal Six operations. Moreover, the runtime elasticities may be established using data that are not subject to clocking error for MODS workhours. Thus, demonstrating that machine runtime does not scale proportionally with sorting volumes makes a facial case for the associated labor usage to exhibit variabilities deviating from the traditional 100 percent variability assumption.

The Public Representative also is mistaken in suggesting that it is inconsistent to apply common data screens to the runtime and workhour models. As noted in the Variability Report, the runtime elasticities are relatively insensitive to data screening, as the runtime and volume data are both generated from reliable machine statistics. The primary study results are the workhour elasticities to be used as volume-variability factors. Applying the same screening criteria to the runtime models simply allows results from the two sets of models to be compared "apples-to-apples" on common sets of observations.

b. The use of models with lagged workload and seasonal dummy variables is adequately and amply justified

A puzzling claim of the Public Representative is that the choice of models with lagged output and seasonal dummy variables were not "explicitly defend[ed]." PR Comments at 16. In fact, the inclusion of lagged output and seasonal dummy variables are justified both on a priori

considerations and with specification testing. The Variability Report notes that sorting workhours may adjust to volumes on longer time scales than the monthly observations, and specifically notes the use of previous-year (SPLY) data in managing workhours. Additionally, the inclusion of seasonal (monthly) dummy variables is based on graphs of variations in hours and TPF volumes showing high degrees of co-movement among the variables, but some potential for residual seasonal variability in workhours, as well as to allow potential effects of peak-season staffing practices on letter- and flat-sorting workhours. Variability Report at 20.

Post-estimation, the inclusion of lagged output and seasonal dummy variables also are explicitly justified by standard specification test statistics. Table 4 of the Variability Report presents test statistics in which the simpler equation (3) in the report is rejected in favor of the Proposal Six specification described by equation (5) (with lags and dummy variables). In addition, extended test results provided in USPS-RM2020-13-1, file results_lag_seasonal_tests.xlsx show that the exclusion of lags and seasonal variables may be rejected separately as well as jointly.[10]

The Public Representative is also off-base in suggesting that the specification test results are merely symptomatic of multicollinearity in the models. As noted in the response to ChIR No. 2, question 4, the adverse effects of multicollinearity are in check for the Proposal Six models. Particularly given the high correlation among current and lagged MODS workloads, it is not at all a foregone conclusion that the relatively small lagged output coefficients would differ significantly from zero. That many if not most of the coefficients are significant individually as well as jointly is not classically symptomatic of multicollinearity. Finally, as the Postal Service showed, the Proposal Six results are robust to alternative specifications limiting the lags included in the model, a common remedy in cases of high multicollinearity. Id.

c. The use of TPF instead of FHP for sorting operations' output is appropriate, and key findings of the Variability Report are robust to substituting FHP as an output measure

The Public Representative characterizes previous modeling approaches using FHP to measure sorting output as having been "promising" (PR Comments at 17), notwithstanding the Commission having rejected those models as well as the Postal Service's models in Docket No. R2006-1. The Public Representative largely ignores the operational justification for the use of TPF discussed in the Variability Report. In short, all sorts (measured by TPF) cause both machine utilization and labor usage in the activities comprising sorting operations, not just the first sorts that are counted as FHP. Variability Report at 7.

Insofar as FHP is an incomplete measure of number of sorts performed in the operations, the reasons for preferring TPF to TPH articulated in the response to ChIR No. 3, question 1(a) apply to FHP a fortiori. FHP is, moreover, a less-homogeneous output measure than TPF (or TPH) in that it does not capture differences in the amount of sorting required to finalize the mail due to variations in the presort mix that plants may encounter. As a simple example, consider the

[10] While the lagged outputs are not jointly significant in the AFSM 100 model with the Proposal Six sample period and screens, the first lag is individually significant.

processing of 3-digit and 5-digit presort letters. The accepted cost avoidance models for letters and flats show that the latter has lower cost because 5-digit presort avoids at least one incoming (primary, SCF, or managed mail) sort required for the 3-digit presort piece. The avoided sort(s) would be recognized by the TPF and TPH MODS volume measures, but not FHP where each piece would get a single FHP count in principle.

Nevertheless, FHP exhibits relatively high correlations with other MODS volume measures and, as noted above, the Postal Service's measurement of FHP for automated sorting operations has been improved since Docket No. R2006-1. It is, of course, possible to let the data speak, as FHP data are included in the dataset; the Public Representative chose not to avail himself of the opportunity to test whether the use of FHP for mail processing variability estimation remains promising. The table below shows TPF and FHP elasticities for workhours from models estimated including both output measures in the models for the AFSM 100 and DBCS groups.[11]

Table 3. Workhour elasticities from models using both FHP and TPF as output variables (FY2016-2019 sample period)

Elasticity	AFSM 100	DBCS
FHP	-0.031	0.033
Std. Error	0.078	0.069
TPF	0.806	0.957
Std. Error	0.120	0.072
Total	0.776	0.989
Std. Error	0.091	0.036
Proposal Six	0.774	0.976
Std. Error	0.091	0.032

Source: USPS-RM2020-13-5, results_seasonal_fhp2.xlsx and analysis_seasonal_fhp2.txt

Table 3 shows that the multiple-output model unequivocally favors the use of TPF as the output measure determining sorting labor. The total elasticities with respect to all outputs in these models do not differ materially from the Proposal Six estimates, strongly implying that any dimensions of sorting output that might be captured by FHP but not TPF do not account for the lower flat sorting elasticities. In addition, the coefficients on the FHP variables are statistically insignificant, and the combined FHP elasticities are small and not significantly different from zero.

[11] FHP measurement for FSS operations is nonstandard, with FHP and TPF identical. Thus, FHP and TPF elasticities for the FSS operation group are identical and not reported separately here. The FSS, unlike other equipment, records a single piece handling (TPF or TPH) representing both sort passes in the two-pass sequencing process. FSS TPF therefore measures the unique number of pieces handled in the FSS operation, but does not subtract pieces that may have been handled in upstream AFSM 100 operations.

Models using only FHP also do not alter the general picture that elasticities for the flat operation groups are lower than that for the DBCS operation group.[12] Table 4 shows elasticities for FHP models using both ordinary least squares and instrumental variables estimation.[13]

Table 4. Workhour elasticities from FHP models, OLS and IV estimation (FY2016-2019 sample period)

Elasticity	AFSM 100	DBCS
Fixed Effects-OLS	0.679	0.833
Std. Error	0.110	0.075
Fixed Effects-IV	0.685	0.857
Std. Error	0.114	0.072

Source: USPS-RM2020-13-5, results_seasonal_fhp.xlsx and results_seasonal_fhp_iv.xlsx

Table 4 shows a similar relative difference between the DBCS and AFSM 100 elasticities as the Proposal Six results, though the point estimates themselves are lower in both cases. The FHP-based elasticities have substantially higher estimated standard errors than the Proposal Six elasticities, such that the Proposal Six AFSM 100 and DBCS elasticity estimates are within the 95 percent confidence intervals for the FHP-based estimates using both OLS and IV models. The IV estimates show little difference from the OLS estimates, suggesting that current FHP methodology for automated sorting operations does not lead to significant measurement error. While FHP may not be theoretically or empirically preferred to TPF as a sorting operation output measure, the data do appear to be relatively reliable.

d. The Public Representative's analysis of the productivity screening grossly misrepresents the Proposal Six procedures and neglects the underlying econometric issues for data screening

The Public Representative criticizes the productivity screen as "results driven" and as not considering whether excluded data are "truly erroneous." In support of his assertions, the Public Representative provides an analysis purporting to show that the Proposal Six screening procedures "wrongly deleted" most of the observations subject to the screens (PR Comments at 22-23). However, the Public Representative's deleted observation analysis is founded on an egregious error in identifying the set of observations covered by the screens and should be disregarded. The Public Representative otherwise fails to articulate any feature of the productivity screen that violates appropriate econometric practice.

[12] The FSS OLS elasticity using FHP is the same as the elasticity using TPF, as noted above. The IV estimate is statistically similar to the AFSM 100 IV estimate. See USPS-RM2020-13-5, file analysis_seasonal_fhp_iv.txt.

[13] The IV models use Prof. Roberts's identification scheme, using DBCS FHP to instrument for AFSM 100 FHP, and AFSM 100 FHP to instrument for DBCS FHP. Previous testing showed these instruments were plausibly exogenous and passed weak-instrument tests, which is generally considered sufficient for workable instrumental variables.

As is noted in the Variability Report as well as the Public Representative's discussion, the screening criterion is TPF per workhour (a labor productivity measure). Variability Report at 21. However, the 'Screens Rev 11.27.sas' program in the Public Representative's library reference selects observations based on *workhours* rather than productivity.[14] The Public Representative's analysis then proceeds to tabulate "wrongly deleted" observations using productivity criteria without consideration of the actual bounds of the five percent productivity screen, which are shown in Table 1 of the Variability Report. Variability Report at 21. As a result, the Public Representative counts as "wrongly deleted" not only some observations that are unusable (e.g., because TPF, hours, or both are zero), but also a substantial number of observations that actually pass the productivity screen and are *not* deleted from the regression samples.[15]

Table 5, below, accounts for observations from the Public Representative's analysis that are usable and observations within the range of the 5% screens for the DBCS group. In the Public Representative's 1% tails analysis, a substantial fraction of the observations he extracts are unusable due to zero workhours or workload, and 57% of the usable observations fail the productivity screen. However, none of the 28 observations the Public Representative identifies as "wrongly deleted" are actually deleted, since his criterion for identifying them was to select observations in the range of 6,400-7,800 TPF/hour productivity—within the range of productivities admitted by the 5% screen (see also Table 6, below).

Overall, the 5% screen retains 43% of the otherwise usable observations from the 1% tails of the workhour distribution. For observations with less extreme values of workhours, in the Public Representative's 5% and 10% tails data, the Proposal Six screens actually retain most of the usable data—67% from the 5% tails and 76% from the 10% tails. In the 5% and 10% scenarios, a majority of the observations the Public Representative counts as "wrongly deleted" were not deleted at all, particularly in the 5% and 10% tails (of workhours) data the Public Representative analyzes.[16] The Proposal Six screens fundamentally do not eliminate observations solely for taking on relatively extreme values of workhours.

---

[14] E.g., the statement identifying DBCS observations subject to the 5% tails screen is "if Hours le 1283.2815 | Hours ge 42112.0105 then output db595;". Compare Table 1 of the Variability Report (at 21) for the productivity cutoffs.

[15] The Public Representative's calculations also inconsistently treat observations as "wrongly deleted" between the 1%, 5% and 10% scenarios. As shown in Table 5, the number of observations not identified as wrongly deleted by the Public Representative drops between the 1% and 5/10% scenarios. Since the implied residual of not "wrongly deleted" observations in the 5% and 10% scenarios is the same as the "wrongly deleted" observation count in the 1% scenario, the likely cause is an arithmetic error on the Public Representative's part.

[16] Likewise, for AFSM 100 and FSS operations, the productivity screen retains most of the usable observations from the outer tails of the distributions of workhours. See USPS-RM2020-13-NP4, Tholds and Deleted Obs final rev 11.27.check.xlsx.

Table 5. Tabulation of DBCS observations in PR screening analysis

Row	Category	1% Tails	5% Tails	10% Tails
1	Observations in PR analysis	183	941	1907
2	PR "Wrongly Deleted"	28	913	1879
3	PR Not "Wrongly Deleted"	155	28	28
4	Usable Observations (TPF>0 & Hours>0)	106	809	1775
5	% of Obs. Usable	58%	86%	93%
6	Obs. Not Deleted in Proposal Six (Pass 5% Productivity Screen)	46	544	1341
7	Obs. Deleted (Fail Productivity Screen)	60	265	434
8	% of Usable Obs. Deleted	57%	33%	24%

Sources:
(1)-(2): USPS-RM2020-13-NP4, Tholds and Deleted Obs final rev 11.27.check.xlsx
(3): R1-R2
(4): Count of obs. with Hours>0 & TPF>0
(5): R4/R1
(6): Count of Obs. Within 5% Productivity Tails (boundaries from Variability Report Table 1), Tholds and Deleted Obs final rev 11.27.check.xlsx. Includes all 28 "Wrongly Deleted" observations in PR 1% analysis.
(7): R4-R6
(8): R7/R4

Beyond the errors in characterizing the effects of the data screens, the Public Representative's discussion of screening does not establish any valid reason for the Commission to conclude that the data screening is methodologically inappropriate. As a theoretical matter, the consequences of including erroneous data are generally more significant than those of not including all potentially valid data. Standard regression assumptions require only that the data included in the regression sample be consistent with the model being estimated, not that all possible observations be included *per se*. The rationale for removing erroneous observations is that the process causing the data errors typically is not part of the model. Otherwise, screening procedures tend to be situation-specific in line with the data quality issues for particular applications. The key requirement is to avoid clearly inappropriate methods such as screening directly on the dependent variable of a regression.[17]

[17] Such a screen would truncate the distribution of the dependent variable and thus violate standard distributional assumptions for regression modeling. The Proposal Six models do not screen on workhours for precisely this reason. See also Frank R. Hampel, et. al., *Robust Statistics*, John Wiley & Sons, New York, 1986, p. 70 ("Any way of treating [i.e., rejecting] outliers which is not totally inappropriate, prevents the worst"); R. Dennis Cook and Sanford Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, New York, 1982, p. 104 ("If the influential cases correspond to gross measurement errors, recording or keypunching errors, or inappropriate experimental conditions, then they should be deleted or, if possible, corrected").

Given that the Public Representative counts as "wrongly deleted" observations that are within the bounds of the productivity screen based on the 5 percent tails, there does not appear to be any controversy regarding the presumptive validity of the included observations under the 5 percent tails screen. Nor is there any apparent disagreement that screening at the ten percent tails is unnecessarily restrictive. It seems inconsistent for the Public Representative to simultaneously harbor concerns about hidden errors in the data while seemingly militating for the inclusion of observations that are potentially anomalous. In any case, the effective dispute is over how far in the outer tails any cutoffs should be. Table 6, below, expands the Variability Report's Table 1 to include the 1% and 10% cutoffs.

Table 6. Productivity screen cutoffs, 1%, 5%, and 10% tails

Operation	Lower 1% Cutoff	Lower 5% Cutoff	Lower 10% Cutoff	Median	Upper 10% Cutoff	Upper 5% Cutoff	Upper 1% Cutoff
AFSM 100	531	733	836	1,225	1,686	1,855	2,312
DBCS	5,355	6,299	6,683	8,314	10,508	11,219	13,502
FSS	465	548	592	787	1,048	1,184	2,348

Source: USPS-RM2020-13-1, analysis.txt

Percent of median

	Lower 1% Cutoff	Lower 5% Cutoff	Lower 10% Cutoff	Median	Upper 10% Cutoff	Upper 5% Cutoff	Upper 1% Cutoff
AFSM 100	43%	60%	68%	100%	138%	151%	189%
DBCS	64%	76%	80%	100%	126%	135%	162%
FSS	59%	70%	75%	100%	133%	150%	298%

Even at the 1% tails, it is perhaps notable that none of the values in Table 6 are clearly erroneous, particularly for DBCS operations. Observations for AFSM 100 operations less than half the median or FSS observations more than twice the median may both be regarded as at least being anomalous, and the 5% tails screen would exclude such values. However, if the Commission prefers a more inclusive screen, it may adopt the less restrictive 1% screen, the results of which (in folder USPS-RM2020-13-1, results_seasonal.xlsx) are qualitatively and statistically similar to the 5% screen results used for Proposal Six.

e. The Public Representative's claim that 'incidental allied labor' is handled inconsistently between letter and flat operation groups is incorrect

The Public Representative contends that flat preparation operations should be removed from the flat operation groups to improve consistency of the treatment of "incidental allied labor" across the DBCS and flat operation groups. PR Comments at 24. The Public Representative's contention stems from two main errors. First, he incorrectly equates flat preparation

operations with "incidental allied labor" as the term is used in the Variability Report. Second, and more significantly, he is wrong to suggest that eliminating flat preparation operations would improve the conceptual consistency of the operation groups, both between letters and flats, but also especially within the flat groups.

The Public Representative's equation of flat preparation with incidental allied labor appears to over-read a statement in the Variability Report that the incidental allied labor component of distribution activities "would also include" flat preparation as a matter of classification. Variability Report at 8. However, incidental allied labor also includes, generally, "activities such as moving mail and equipment into and out of the operations carried out by employees clocked into the sorting operation." Id. Such activities are not flat-specific, and occur in both letter and flat operations.

The Public Representative also neglects to properly consider the role of flat preparation operations in activities related to feeding mail into automated flat sorters. Including flat preparation in the flat groups serves to improve consistency across flat machine types and between DBCS and flats by ensuring all work related to feeding mail is included in the operation groups. Work related to feeding mail in automated flat operations generally is divided among distribution and flat preparation operations, whereas it is carried out within the distribution operations for DBCS. Thus, there is also some overlap between flat preparation and the "runtime" activities described in the Variability Report at 6.

A DBCS operator feeds the machine by removing pieces of mail from letter trays and placing them on an input stacker for feeding; this work would be included in the MODS DBCS operation for the employee. Similarly, AFSM 100 clerks operating feed stations for machines without automated induction (AI) remove pieces from flat tubs or flat mail carts and place them on input stackers; this labor would be included in the LDC 12 AFSM 100 distribution operation. Loading the flat mail carts with bundled pieces is carried out in the 035 flat prep operation.[18] For AFSM 100s with AI, all activities for loading the machine are carried out within the AI prep operation 140, where employees place flats to be processed into automation-compatible trays that circulate to and from the AFSM 100-AI machines' feeders; the feed clerk's responsibility is to monitor the automated feed stations. Some of the operation 140 work would be carried out in operation 035 for non-AI machines, but some would be part of the AFSM distribution operations. FSS is analogous to AFSM 100 with AI. See also Variability Report at 6.

The combined set of distribution and (for flat groups) flat preparation operations encompasses the entirety of the interrelated activities related to the automated sorting processes and provides a more consistent treatment of operations, providing much fuller coverage of the corresponding cost pools by the variability models. Finally, it may be noted that the treatment

[18] The situation unique to flats is the need to remove and face pieces from bundles for piece distribution—where bundle-based mail preparation is central to commercial flats but not letters. The flat preparation operations (035, 140, and 530) do not have letter analogues.

of flat preparation in Proposal Six is consistent with the Commission's accepted treatment of flat preparation as components of the Cost Segment 3.1 cost pools for AFSM 100 and FSS operations. See, e.g., Docket No. RM2011-12, Order No. 920 at 8-9 (esp. footnote 13); Docket No. RM2018-10, Order No. 4855 at 17.

IV. Conclusion

As described above, the Public Representative's criticisms of the Proposal Six methodology rely critically on mischaracterizations of the operations covered by the analysis and faulty analysis of the Proposal Six datasets. Many of his other critiques, such as those claiming a lack of justification for aspects of the regression specifications, are amply addressed in the Variability Report or Postal Service responses to Chairman's Information Requests. The Public Representative's comments are also notable in what they do not show: any substantial reason why the Proposal Six elasticities exhibit statistical bias or inconsistency, let alone a source of bias that is specific to the flat operations groups. Nor does the Public Representative show that the Proposal Six analysis would not improve the quality of cost data over current methodologies. Establishing an empirical basis for mail processing variabilities, whether or not consistent with the results of the assumptions underlying current methodology, is a significant improvement meriting the Commission's careful consideration.